1.

Bayes' Theorem: $P(A|B) = \frac{P(B|A)*P(A)}{P(A)*P(B|A) + P(\neg A)*P(B|\neg A)}$

Density Function (NRV): $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$

P(A) = .8

P(B|A) = $\frac{1}{\sqrt{2*\pi*36}} * e^{\frac{-(4-10)^2}{2*36}}$ = .040328

P($\neg$A) = .2

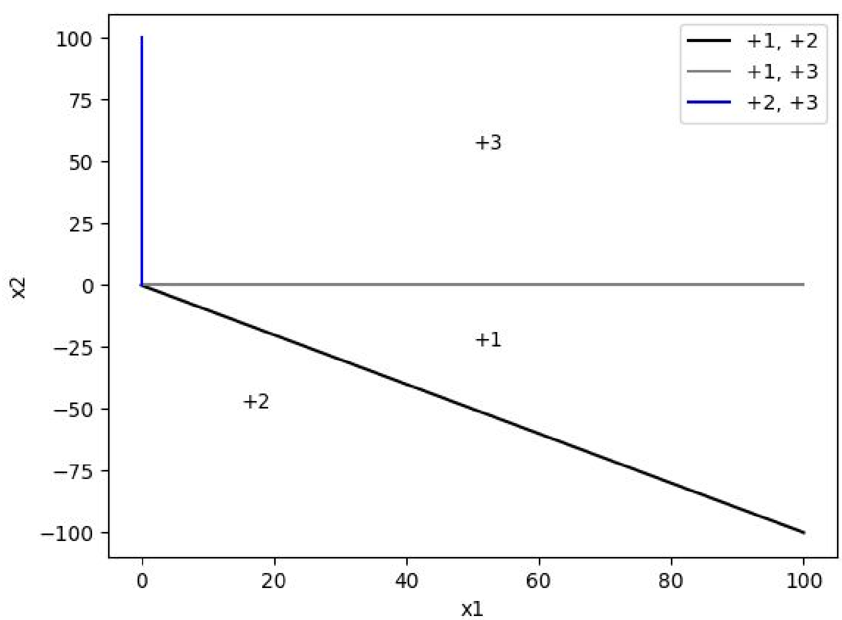
P(B|$\neg$A) = $\frac{1}{\sqrt{2*\pi*36}} * e^{\frac{-(4-0)^2}{2*36}}$ = .053241

Equation: $\frac{.8 * .040328}{.8 * .040328 + .2 * .053241}$

Answer: .751852 ~ 75.2%

2.

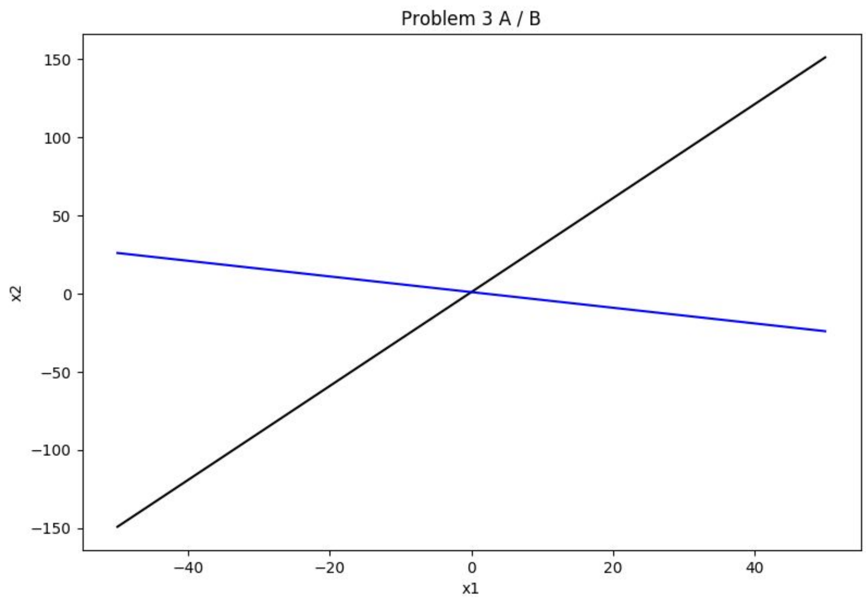
a. $0 = x_2 + x_1$
b. $0 = x_2$
c. $0 = x_1$
d.



e. The decision boundaries would be quadratic because QDA enables different feature covariance matrices. LDA assures that the decision boundaries are linear.
f. The assumption feature independence that comes with using naive Bayes could skew the probability estimate.

3.

a / b



Black: $1 + 3X_1 - X_2 = 0$

Blue: $-2 + X_1 + 2X_2 = 0$

Above $X_2 == 0$

Black points < 0

Blue points > 0

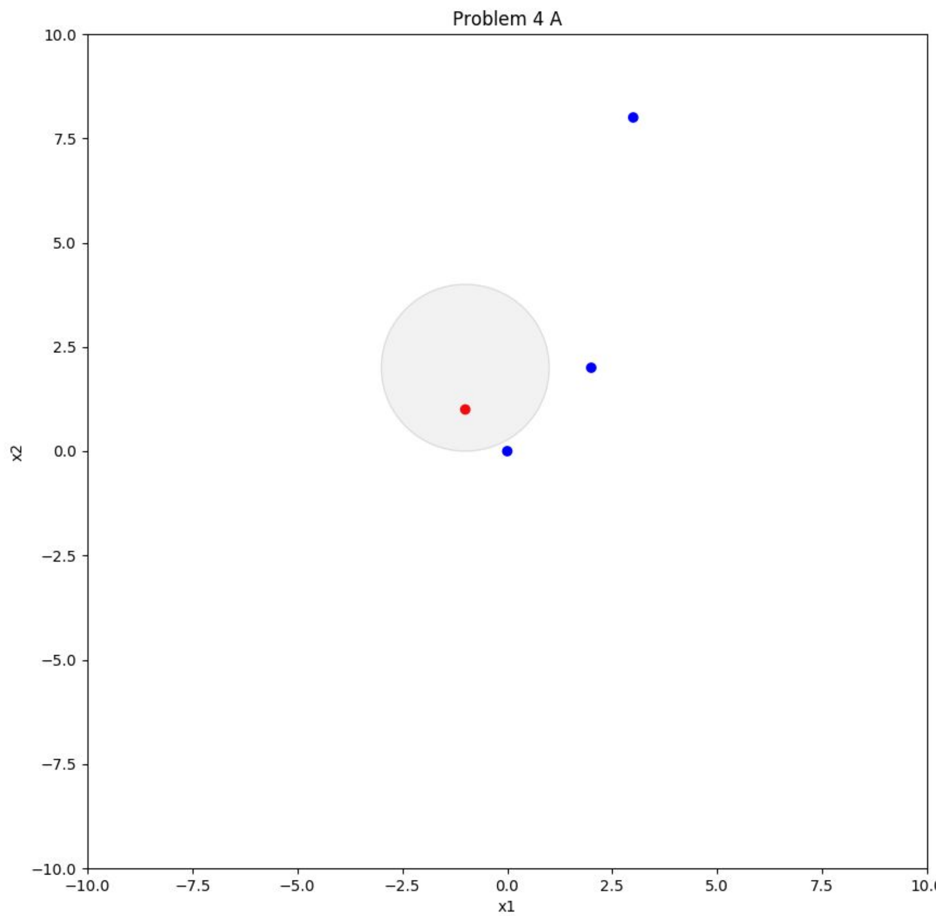
Below $X_2 == 0$

Black points > 0

Blue points < 0

4.

a.



b.

inside circle <= 4

outside circle > 4

c. shown on graph

d. Foiling out the two non-linear parts results in a linear equation in terms of $X_1,\ X_1{}^2, X_2, X_2{}^2$.

$(1 + X_1) * (1 + X_1) + (2 - X_2) * (2 - X_2) - 4 = 0 \ \rightarrow$

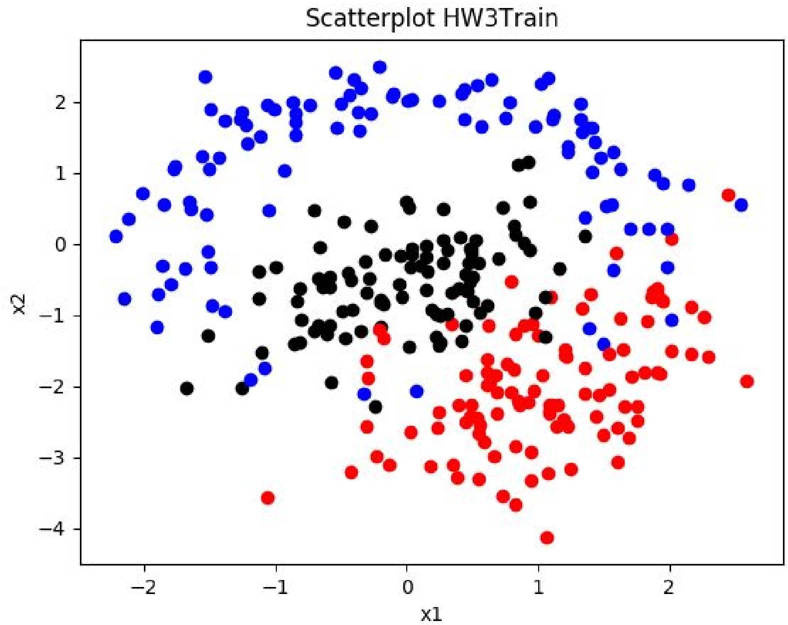
$1 + 2X_1 + X_1{}^2 + 4 - 4X_2 + X_2{}^2 - 4 = 0 \rightarrow$

$1 + 2X_1 - 4X_2 + X_1{}^2 + X_2{}^2 = 0$

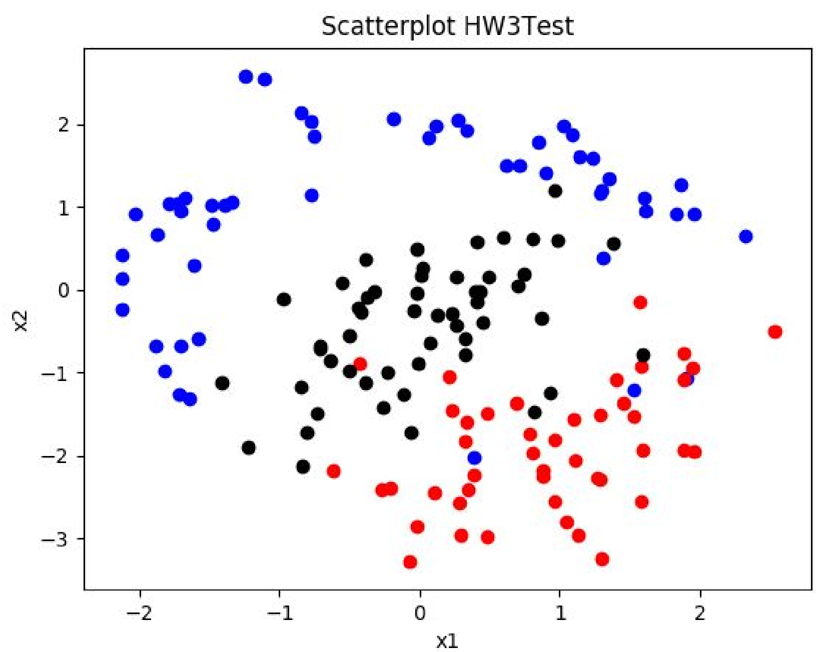
5.

a.

i.

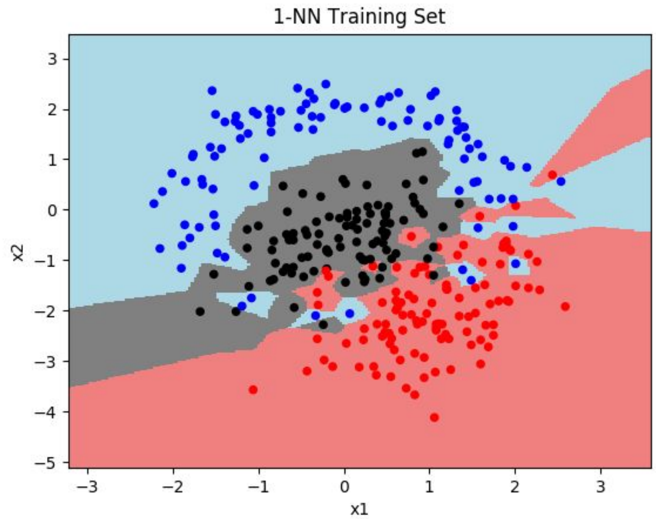

Scatterplot HW3Train
x2
x1
2
1
0
−1
−2
−3
−4
−2
−1
0
1
2

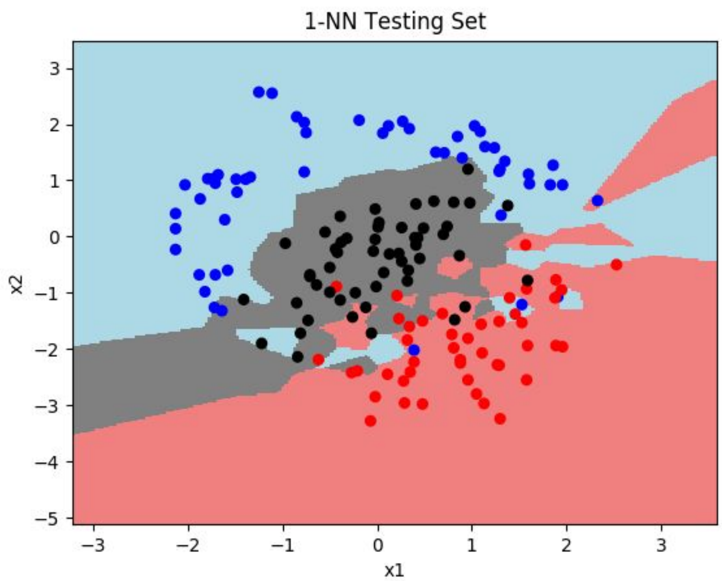
ii.


Scatterplot HW3Test
x2
x1
2
1
0
−1
−2
−3
−2
−1
0
1
2

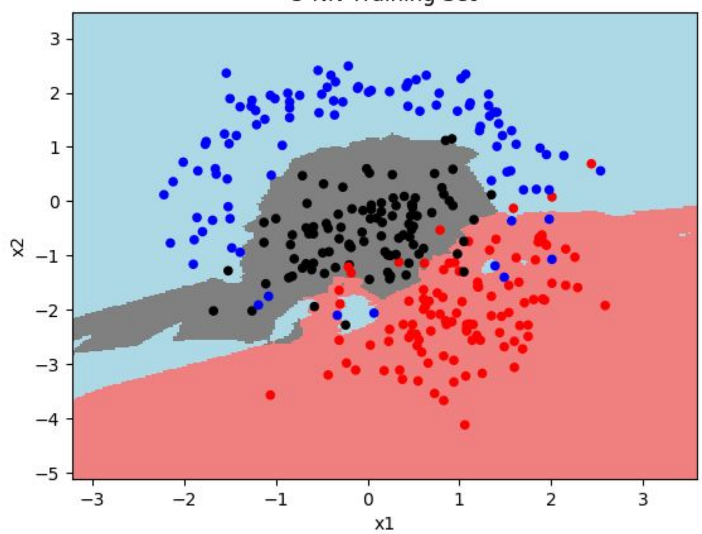
b.

i.

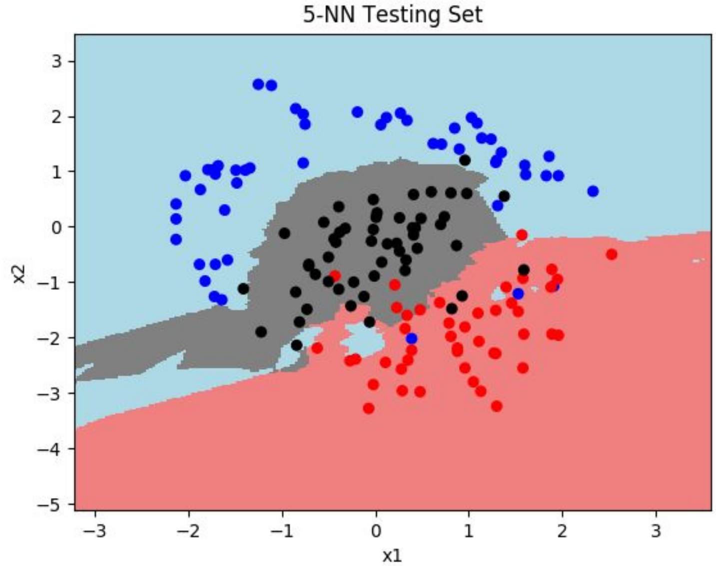

1-NN Training Set
x1
x2

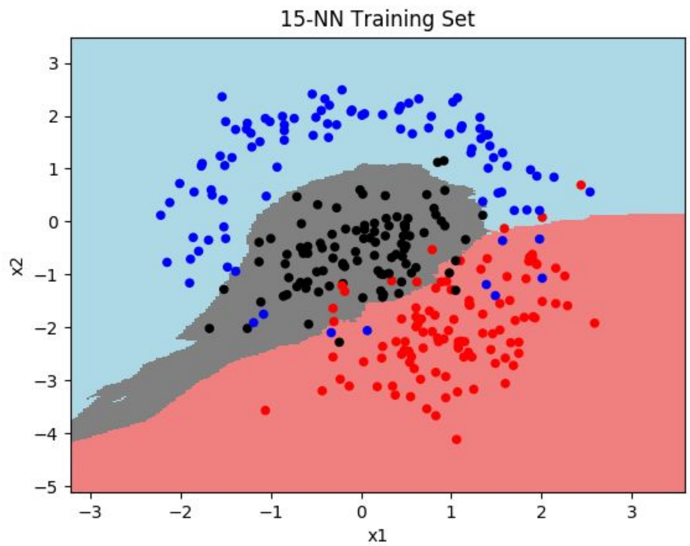

1-NN Testing Set
x1
x2

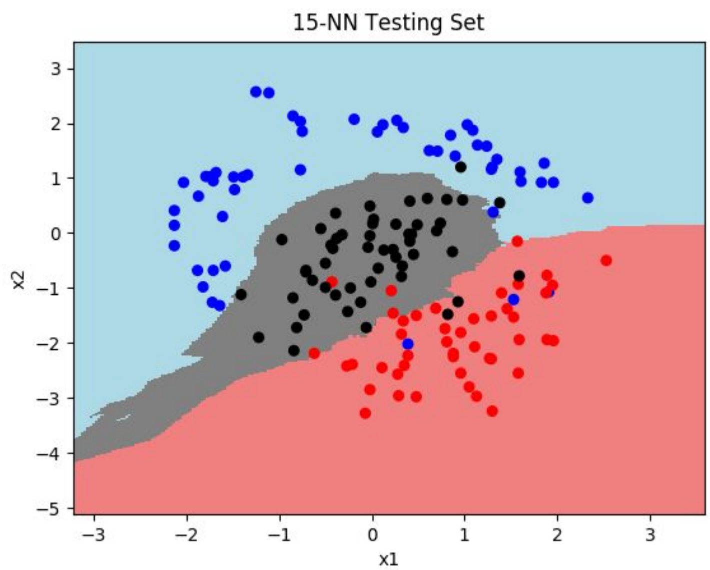
5-NN Training Set



5-NN Testing Set



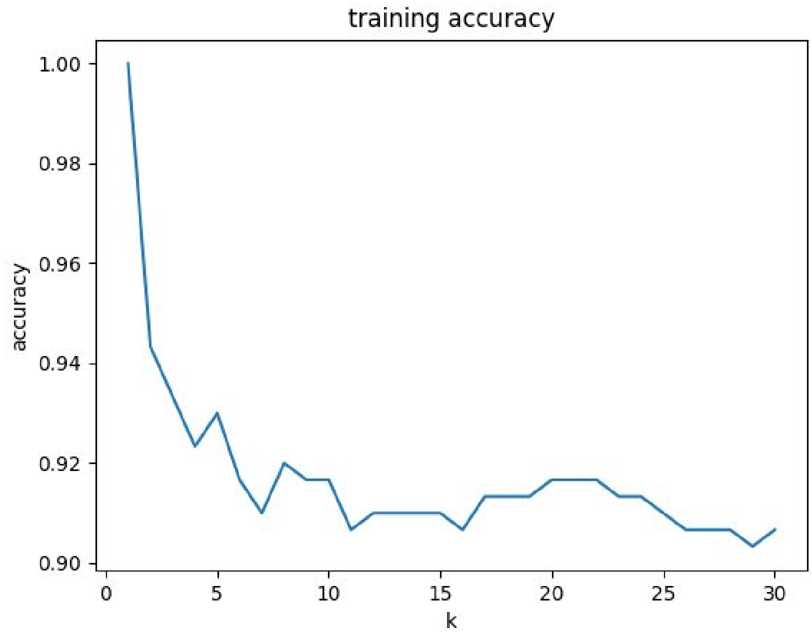




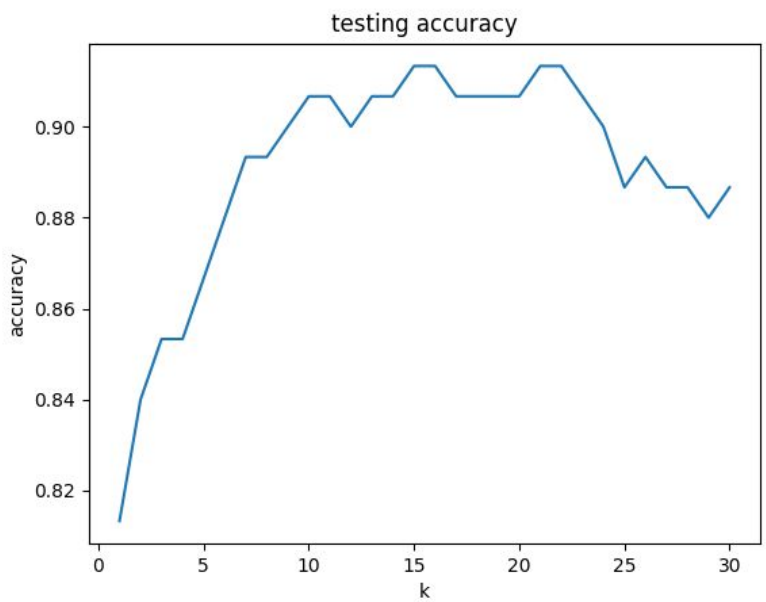
ii.

Same: shape and relative position of regions.
Different: as k increases the borders of each region are easier to trace and become smooth. Fewer pockets of color x completely surrounded by color y as k increases.

iii.



iv.

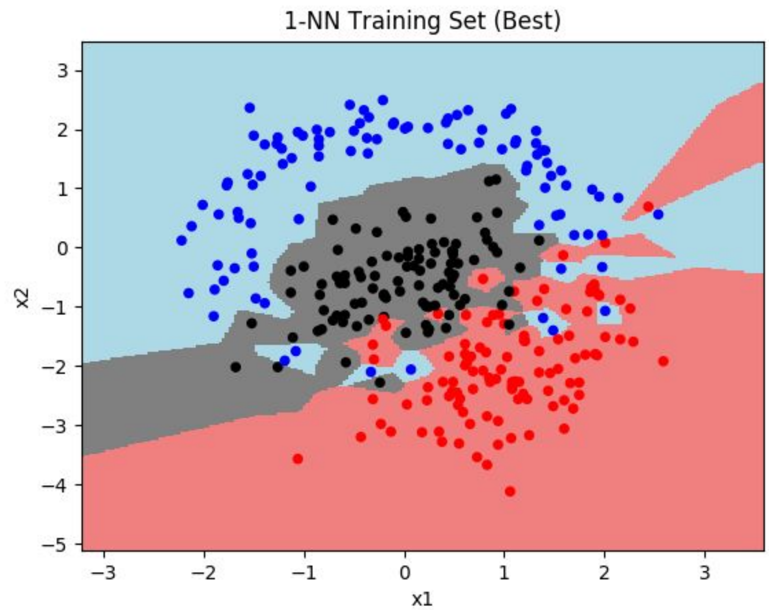


v.

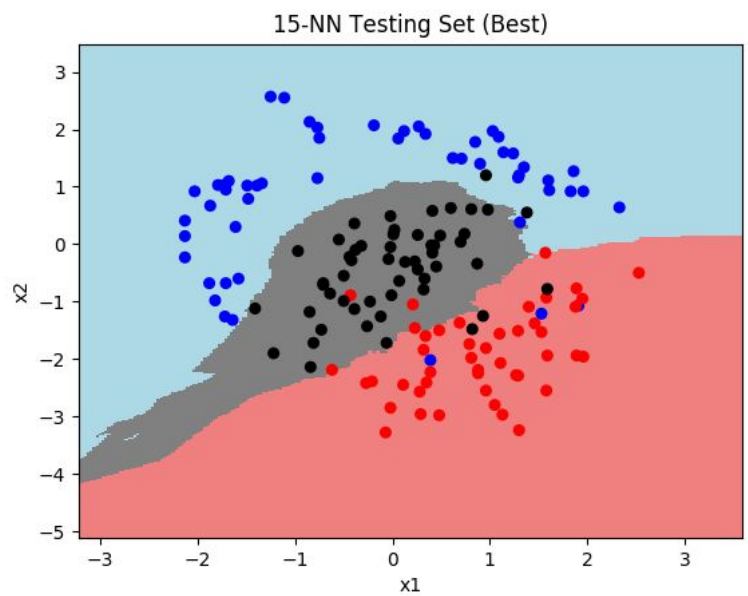
Train: k=1, 100%

Test: k=15, 91.3333%

vi.



vii.



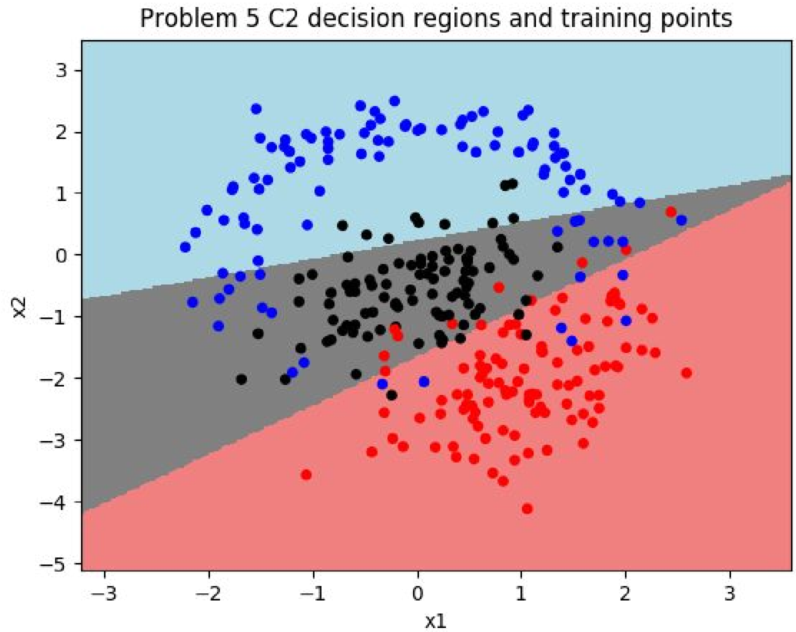
c.

i.

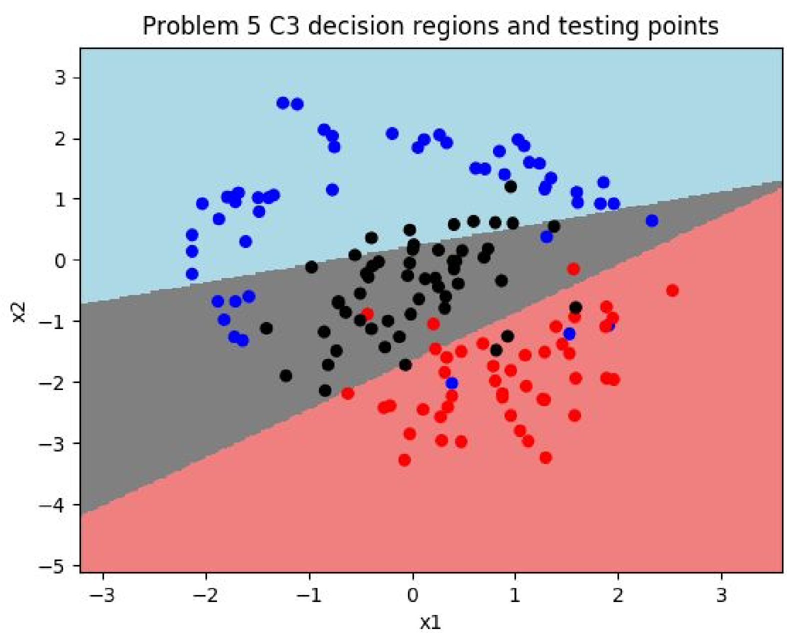
Train: 0.8334

Test: 0.81334

ii.



iii.



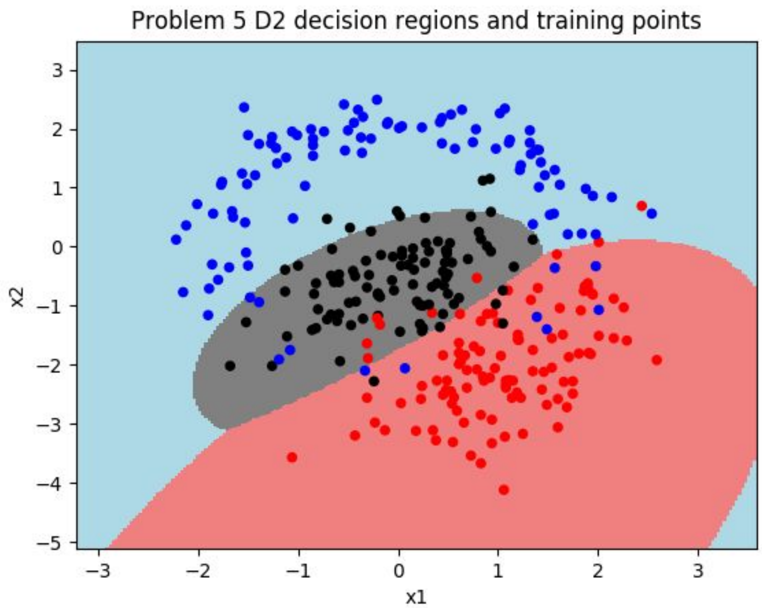
d.

i.

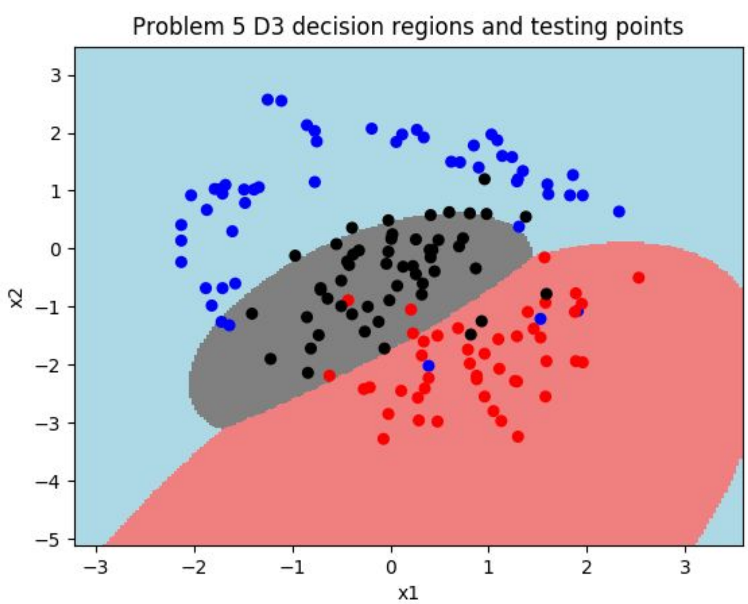
Train: 0.89667

Test: 0.8667

ii.



iii.

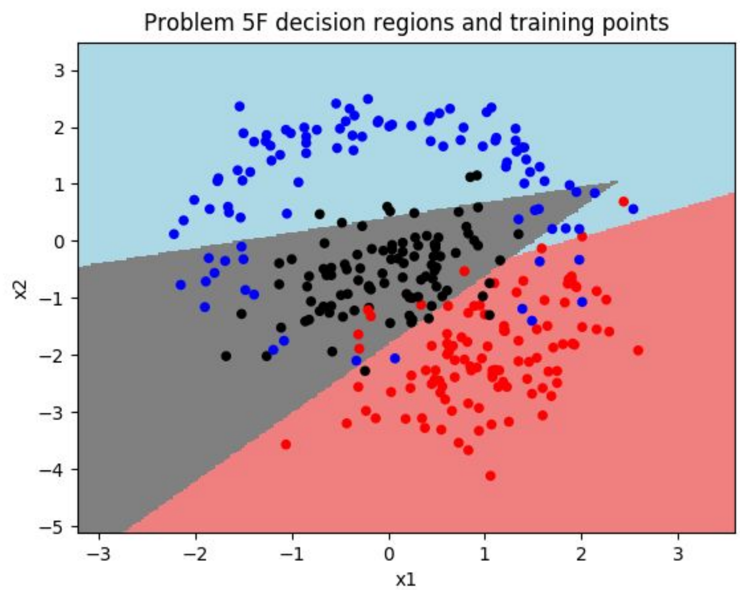


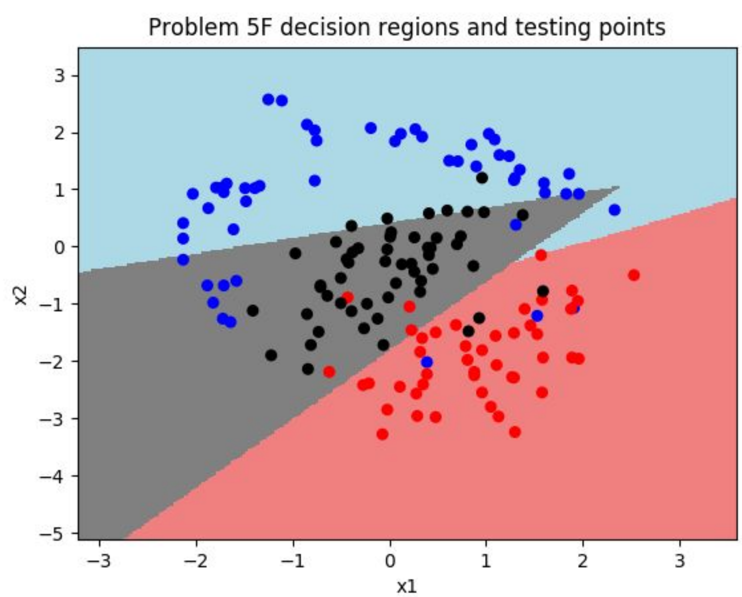
e.

Knn: disjoint borders and the outermost edge of a region doesn't guarantee uniformity inside. No specific shape created.
LDA: Linear borders that clearly define the break between color x and color y. created zones that resemble triangles. The regions are connected.
QDA: Rounded decision borders that resemble ovals. The regions are connected with smooth borders.

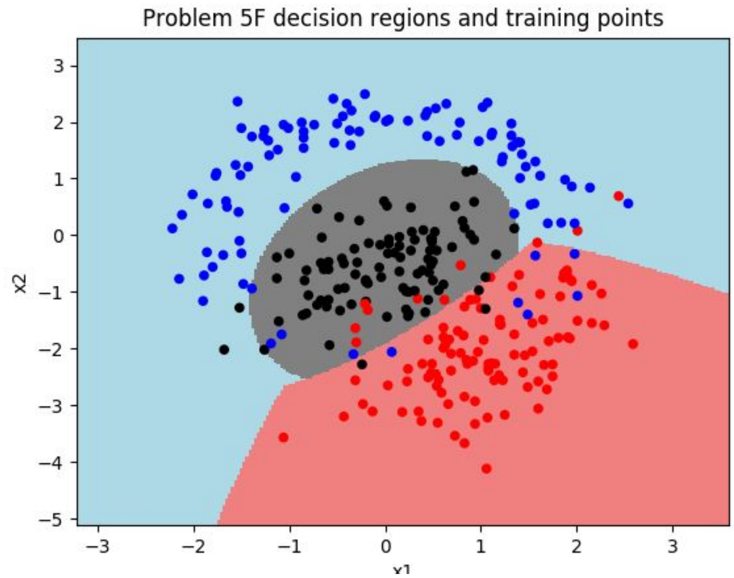
f.

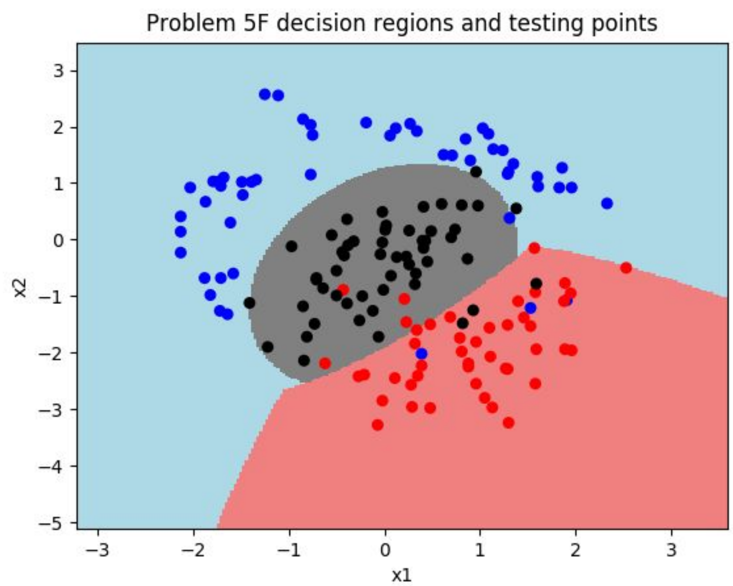
p = 1
c value = 1.0
Train: 0.85
Test: 0.84





p = 2
c value = 1.0
Train: 0.92
Test: 0.91334



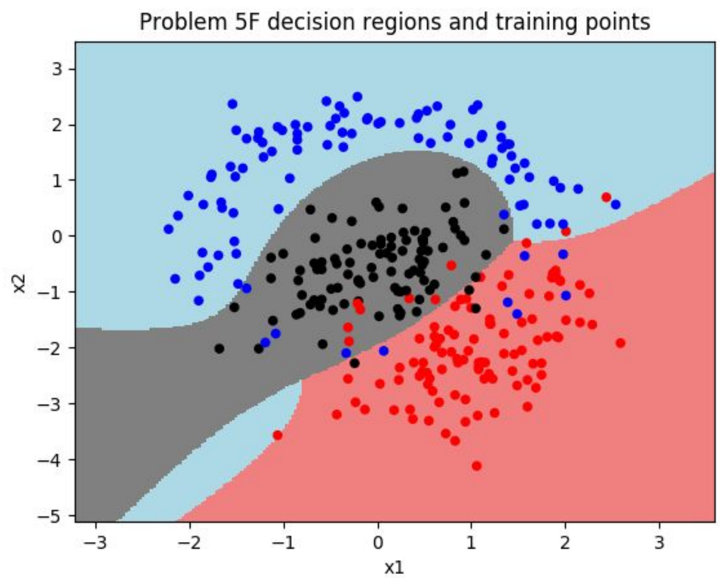


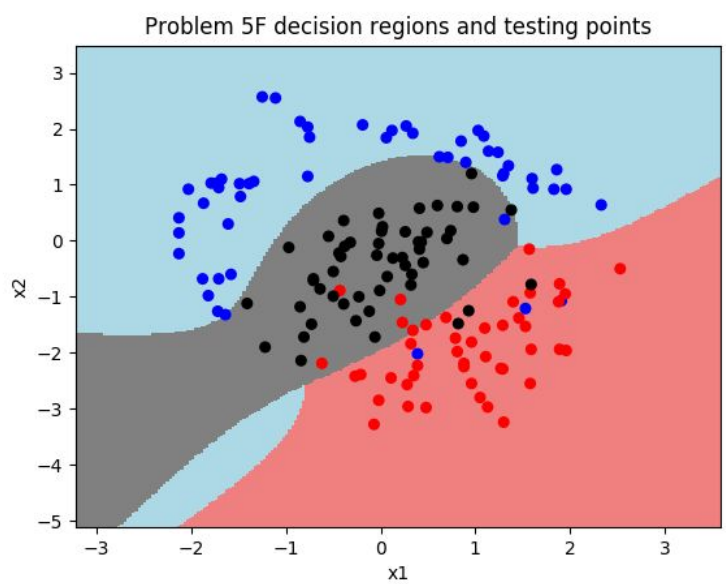
p = 3

c value = 1.0

Train: 0.91667

Test: 0.90667



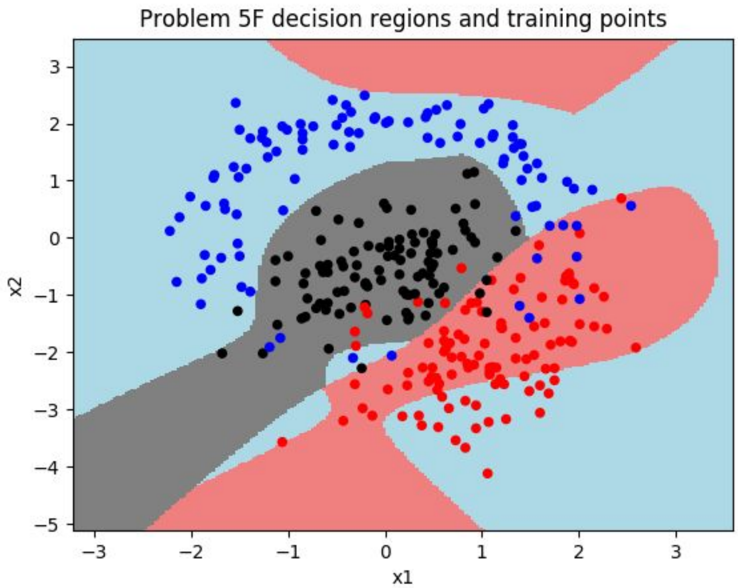


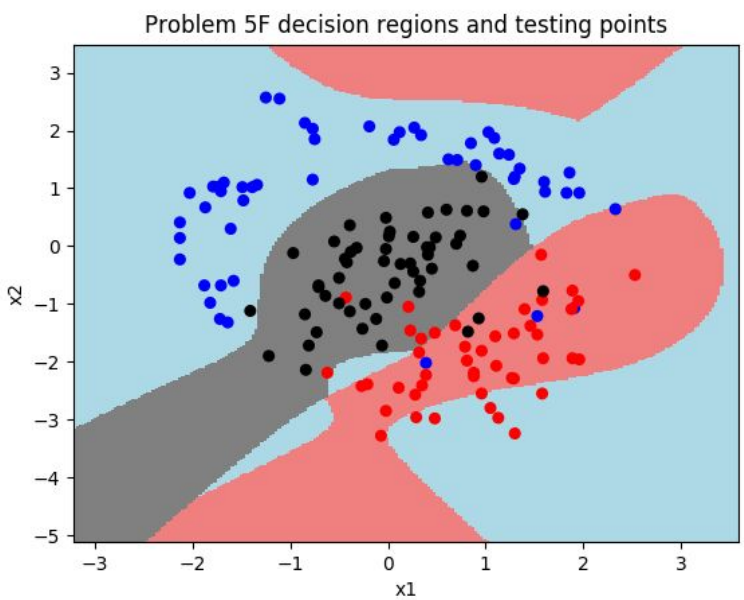
p = 4

c value = 1.0

Train: 0.83667

Test: 0.84667





g.

p: 1
c: 1.0
gamma: 0.01
Train: 0.85667
Test: 0.8334

p: 2

c: 1.0
gamma: 0.01
Train: 0.85667
Test: 0.8334

p: 3
c: 1.0
gamma: 0.01
Train: 0.85667
Test: 0.8334

p: 4
c: 1.0
gamma: 0.01
Train: 0.85667
Test: 0.8334

h. poly performed better than rbf by an average of .06 which is not trivial. SVM performed similar to QDA / Knn and the resulting region shapes mirrored this observation. Both methods resulted in rounded decision region borders that were connected. LDA was an obvious visual outlier as the linear decision region borders were a stark contrast to the rounded or abstract regions created by the other 3 methods.

i. If the data was truly randomly generated Knn would likely perform the best because the other 3 methods are heavily reliant on the data implicity grouping itself.